

Listed are abstracts from recent papers by IBM authors. Inquiries should be directed to the publications cited.

Analysis and Improvement of Photostore Error Rates, R. T. Chien,* D. T. Tang, E. S. Barrekette, and A. M. Katcher, *Proceedings of the IEEE* **56**, No. 5, 805-812 (May 1968). Statistical experiments have been performed on the photostore to test its reliability as a mass storage system. Errors are analyzed with respect to rate, distribution, and susceptibility to correction by several codes. The experimental results reported are based on disks recorded by optical techniques.

* University of Illinois, Urbana, Illinois.

Abstracts

Analyzing Errors with the Boolean Difference, F. F. Sellers, Jr., M. Y. Hsiao, and L. W. Bearnson, *IEEE Transactions on Computers* **C-17**, No. 7, 676-683 (July 1968). The Boolean difference is defined. It is shown through example how the Boolean difference is used to analyze the effect of errors on the outputs of logic circuits. Examples are given of error-detection problems, analysis of redundant logic, and the generation of diagnostic sequences.

An Autonomous Reading Machine, R. G. Casey and G. Nagy, *IEEE Transactions on Computers* **C-17**, No. 5, 492-503 (May 1968). An unconventional approach to character recognition is developed. The resulting system is based solely on the statistical properties of the language, therefore it can read printed text with no previous training or a priori information about the structure of the characters. The known letter-pair frequencies of the language are used to identify the printed symbols in the following manner. First, the scanned characters are partitioned into distinct groups of similar patterns by means of a distance measure. Each class (at most 26 are permitted) is assigned an arbitrary label, and an intermediate tape, containing these temporary labels of the symbols in the original sequence, is generated. In the second phase of the program, the matrix of bigram frequencies of the labels is compared to a frequency matrix obtained from a large sample of English text. The labels are then assigned alphabetic symbols in such a way that the correspondence between the two matrices is maximized. The method is tested on a 100 000-character data set comprising four markedly different fonts.

Design of an Asynchronous Main Storage-Central Processing Unit Interface, D. E. Waldecker, *Computer Design* **7**, No. 6, 60-63 (June 1968). This article explains the use of an asynchronous interface between the Main Storage (MS) and the Central Processor Unit (CPU) in a parallel digital computer. The described interface is used in the IBM System/4 Pi family of computers. An asynchronous interface permits the MS characteristics to be changed at any time with no impact to the CPU logic. Thus the MS characteristics may be adjusted to compensate for remote MS location. Additional advantages are: a reduction in logic required for the CPU instruction sequencing control, simplified MS cycle stealing input/output design, and the ability to substitute either faster or slower main stores than originally designed for.

Error Detection and Correction in a Photo-Digital Storage System, I. B. Oldham, R. T. Chien,* D. T. Tang, *IBM Journal of Research and Development* **12**, No. 6, 422-430 (November 1968). An error-correction system has been implemented for data stored in the IBM Photo-Digital Storage System. Hardware is used for encoding and error detection, and a processor-controller is used, on a time-sharing basis, for error correction. A Reed-Solomon code is used to obtain a very low error rate in spite of flaws affecting the recorded bits. This approach is applicable to systems which require complex codes and have a data processor available on a time-sharing basis.

* University of Illinois, Urbana, Illinois.

Experimental Studies in Computer-Assisted Correction of Unorthographic Text, E. J. Galli and H. M. Yamada,* *IEEE Transactions on Engineering Writing and Speech* **EWS-11**, No. 2, 75-84 (August 1968). In conjunction with a feasibility study of applying an experimental dictionary and a computer to a proofreading application, investigations into the nature of conversion errors were made and computer programs for correction of unorthographic machine-readable text were developed. The correction programs were tested with a sample of unproofread technical abstracts containing a high concentration of proofreading difficulties. The error-correction program consists of three levels. In the first level, commonly misspelled words are corrected directly through associative dictionary references. The second level treats a number of common misspelling patterns not readily amenable to direct dictionary correction. Error patterns most probably associated with human operation of transcription devices, or with limitations or malfunctions of conversion equipment, are dealt with by the third level. Beyond the first level, the program can apply up to nine error-correction algorithms, depending on the nature of the error. Because of the combinatorial nature of selections of alternative spellings, the correction algorithms must test several hundred candidates per error, on the average. Hence, a study was performed concerning computer design characteristics desirable for increasing the efficiency of correction procedures.

* University of Pennsylvania, Philadelphia, Pennsylvania.

An Experimental Study of Machine Recognition of Hand-Printed Numerals, R. Bakis, N. M. Herbst, and G. Nagy, *IEEE Transactions on Systems Science and Cybernetics* **4**, No. 2, 119-132 (July 1968). The recognition of hand-printed numerals is studied on a broad experimental basis within the constraints imposed by a raster scanner generating binary video patterns, a mixed measurement set, and a statistical decision function. A computer-controlled scanner is used to acquire the characters, to adjust the raster resolution and registration, and to monitor the black-white threshold of the quantizer. The dimensionality of the decision problem is reduced by a hybrid system of measurements. In the measurement design, three types of measurements are generated: a set of "topological" measurements, a set of logical " n -tuples," both designed by hand, and a large set of n -tuples machine generated at random under special constraints. The final set of 100 measurements is selected automatically by a programmed algorithm that attempts to minimize the maximum expected error rate between every character pair. Computer simulation experiments show the effectiveness of the selection procedure, the contribution of the different types of measurements, the effect of the number of measurements selected on recognition, and the desirability of size and shear normalization. The final system is tested on four data sets printed under different degrees of control on the writers. Each data set consists of approximately 10 000 characters. For this comparison, a first-order maximum likelihood function with weights quantized to 100 levels is used. Error versus reject curves are given on several combinations of training and test sets.

File Organization Schemes Based on Finite Geometries, C. T. Abraham, S. P. Ghosh, and D. K. Chaudhari,* *Information and Control* 12, No. 2, 143-164 (February 1968). Some new schemes, possessing certain desirable properties, for organizing records with binary-valued attributes have been defined. It has been shown that it is possible to construct these filing schemes using finite geometries. The search time for a query involving any k attributes for these filing schemes based on finite geometries is very small in comparison with existing filing schemes. Moreover, the search time does not depend on the number of records. The problem of updating is also quite simple.

* University of Ohio, Athens, Ohio.

High-Speed Binary-to-Decimal Conversion, M. S. Schmookler, *IEEE Transactions on Computers* C17, No. 5, 506-508 (May 1968). This note describes several methods of performing fast, efficient, binary-to-decimal conversion. With a modest amount of circuitry, an order of magnitude speed improvement is obtained. This achievement offers a unique advantage to general-purpose computers requiring special hardware to translate between binary and decimal numbering systems.

The IBM 1975 Optical Page Reader, *IBM Journal of Research and Development* 12, No. 5, 346-371 (September 1968).

Part I: System Design by R. B. Hennis. The IBM 1975 Optical Page Reader, specially built for the Social Security Administration, reads over 200 fonts from quarterly employer reports printed by electric and manual typewriters, business machines, and high-speed printers. Since the SSA has no control over the means used by employers to prepare the reports, many variations in print quality are present. This paper discusses the problems involved in planning and developing a system to read these reports and summarizes the design of the specialized video signal processing circuits and the character recognition logic that are used in the system. Two companion papers treat the latter topics in more detail. Also discussed in the paper is a management information system that permitted detailed analysis of experimental data and accelerated the development process.

Part II: Video Thresholding System by M. R. Bartz. An adaptive video thresholding system is used in the Page Reader to minimize recognition failures due to contrast and line width variation. The main threshold operator is a linear function of the average contrast over a specified area. Adjustments to the threshold level are made by circuits that compute the average line width in a character, and by circuits that filter out spatial noise in the vicinity of the character. The different types of print quality distortions predominant in typewriter printing and business machine printing are handled by switching between different sets of threshold operators.

Part III: Recognition Logic Development by D. R. Andrews, A. J. Atrubin, and K. C. Hu. The design approaches which were used to specify feature measurement logic, recognition reference standards, and decision functions for a multifont character recognition system are discussed. The importance of an intuitive approach to design, as opposed to a fully automated approach, is emphasized. The nature of the problem required an intimate interaction between the designers, who investigated complex pattern recognition problems and proposed design alternatives, and the computer, which relieved the designer of routine testing and evaluation of the tentative design.

A New Algorithm for Inner Product, S. Winograd, *IEEE Transactions on Computers* C-17, No. 7, 693-694 (July 1968). In this note we describe a new way of computing the inner product of two vectors. This method cuts down the number of multiplications required when we want to perform a large number of inner products on a smaller set of vectors. In particular, we obtain that the product of two $n \times n$ matrices can be performed using roughly $n^3/2$ multiplications instead of the n^3 multiplications which the regular method necessitates.

Special Issue on Laboratory Automation, *IBM Journal of Research and Development* 13, No. 1, 2-138 (January 1969). This issue comprises seventeen papers on original work in the field of laboratory experimentation aided by digital computers. The various contributions by IBM authors and guest authors include descriptions of functioning systems involving some aspect of data acquisition, open and closed loop control, and real-time data analysis through the use of such devices as an interactive graphic terminal. Among the applications discussed are nuclear physics, x-ray diffraction and fluorescence, spectroscopy, chemical metallurgy, chemistry, biology, and lens testing. The papers not only describe the various approaches to laboratory automation and the operating features of the systems, but also show how the use of computer-instrument combinations, with appropriate interfaces, can change the entire character of laboratory experimentation. *Foreword* by J. D. Swalen; *Computer-operated X-ray Laboratory Equipment* by H. Cole; *Automation of a Wide-range, General-purpose Spectrophotometric System* by P. M. Grant; *New Research Techniques for the Life Sciences* by G. D. McCann; *Computer-assisted Spectroscopy* by B. Johnson, T. Kuga, and H. M. Gladney; *Combination of On-line Analysis with Collection of Multicomponent Spectra in an On-line Computer* by N. P. Wilburn and L. D. Coffin; *An Interactive Graphics System for Nuclear Data Acquisition* by J. Birnbaum, T. Kwap, M. Mikelsons, P. Summers, J. F. Schofield, and F. Carrubba; *Simulation and Experimental Research* by J. J. Byerley and T. Z. Fahidy; *Computer Facilities for the Laboratory* by T. R. Lusebrink and C. H. Sederholm; *Use of a Time-sharing Computer in Nuclear Chemistry* by J. Fryklund and W. Loveland; *Computer-controlled Optical Spectrometer* by D. M. Hannon, D. E. Horne, and K. L. Foster; *Growth of a Laboratory Computer System for Nuclear Physics* by J. F. Mollenauer; *Measuring Optical Transfer Functions of Lenses with the Aid of a Digital Computer* by J. B. Davis and H. H. Herd; *The Use of Computers at CERN* by R. T. Bell and H. Øverås; *An Experimental System for Time-shared On-line Data Acquisition* by H. A. Reich; *Real-time Reduction of Nuclear Physics Data* by P. R. Bevington; *The Use of a Control Computer in a Chemistry Department* by Y. Okaya; *Use of a Terminal System for Data Acquisition* by K. Konnerth.

State of the Art in Pattern Recognition, G. Nagy, *Proceedings of the IEEE* 56, No. 5, 836-862 (May 1968). This paper reviews statistical, adaptive, and heuristic techniques used in laboratory investigations of pattern recognition problems. The discussion includes correlation methods, discriminant analysis, maximum likelihood decisions, minimax techniques, perceptron-like algorithms, feature extraction, preprocessing, clustering, and nonsupervised learning. Two-dimensional distributions are used to illustrate the properties of the various procedures. Several experimental projects, representative of prospective applications, are also described.